

Raymond Yoo

Amazon Machine Learning Engineer II

Personal Site + Portfolio: raymondyyoo.com

About Me

Technical Skills: LLMs, LangChain, Spark, SQL, Big Data, ML Ops, AWS SageMaker, ECS, lambda, EMR
Web & Mobile: HTML/CSS/vanilla Javascript, NodeJS, ReactJs, Java Spring
Languages: Python, Java, Javascript, Haskell, C, Swift
Interests: Music, AI, Math, Mental Health, Philosophy, Classics, Deep Reinforcement Learning

Work Experience

Amazon NYC (Mar 2022 - Current)

- Machine Learning Engineer
 - Supporting a 100K+ TPS service that applies negative keyword filtering for search queries for sponsored products on Amazon.com with a < 25ms latency budget
 - Productionizing ML models that applied scientists in the org work on for A/B tests and frequent iterations of parametric experiments
 - Working on adding new model capabilities for our positive keyword matching online inference service which uses lexical and semantic match using kNN on the vector space of ads embeddings for sourcing ads to Amazon.com
 - Using Spark + AWS StepFunctions + AWS EMR (with all CDK infrastructure) for our data pipelines for the indexing jobs to create the actual embedding spaces
- Software Engineer II
 - Developing big data pipelines for ML recommendation services running spark on EMR.
 - Productionizing ML ETLs that applied scientists on the team create using spark SQL, EMR, Redshift, lambda, CDK, and SageMaker.
 - Working on a 100+ API service with a ReactJS frontend that saves account managers thousands of hours a year in creating managed service advertising campaigns.
- Software Engineer I
 - full stack web dev: React, Redux, SQL + NoSQL, AWS, Redshift, EMR, data pipelines, Java backend
 - CI/CD using internal tools, git, REST, Agile

Amazon NYC (Summer 2021)

- Software Engineer Intern
 - Full stack web development + data engineering for decorating data to present trailing quarters of advertiser performance on internal website

Adventure Co. Remote (Summer 2020)

- AI/Web Intern
 - Analyzed audio data both traditionally and after converting audio data into spikes (event-based data) for the use of spiking neural networks for classification
 - Collaborated in developing a webpage game using React, based on music and emojis

Research Experience - Rutgers University

Machine Learning Lab

Advisor: Sungjin Ahn Piscataway, NJ (Dec 2020-Dec 2021)

- Undergraduate Research Assistant - Deep Reinforcement Learning (DRL)
 - Implemented OpenAI's Dreamer model from scratch and presented my learnings to the lab
 - Used Dreamer as a baseline model to solve Atari games

Computational Brain Lab

Advisor: Konstantinos Michmizos Piscataway, NJ (Jan 2020-Dec 2020)

- Undergraduate Research Assistant - ML, Deep Learning, Robotics, Deep Reinforcement Learning (DRL)
 - Developed several state-of-the-art DRL algorithms with spiking actor networks using PyTorch
 - Wrote baselines for measuring the performance of neuromorphic versions of DRL algorithms
 - Worked on debugging deep learning models in PyTorch in a linux environment
 - Accepted publication at CoRL20 + TMLR22 (see below for link to published paper).

Education and Honors

Rutgers University New Brunswick

New Brunswick, NJ (2018-Present)

- GPA: 3.73
- **Publications:** <https://arxiv.org/pdf/2010.09635.pdf> + <https://openreview.net/pdf?id=ZPBJPGX3Bz>